# On A Generalized Baillon–Haddad Theorem for Convex Functions on Hilbert Space*

Charles L. Byrne†

November 24, 2014

**Abstract**

The Baillon–Haddad Theorem asserts that, if the gradient operator of a convex and Fréchet differentiable function on a Hilbert space is nonexpansive, then it is firmly nonexpansive. This theorem plays an important role in iterative optimization. In this note we present a short, elementary proof of a generalization of the Baillon–Haddad Theorem.

**Key Words:** Bregman distance, convex function, firmly nonexpansive, gradient, nonexpansive, Baillon–Haddad Theorem, Krasnosel'skii–Mann Theorem.

**2000 Mathematics Subject Classification:** Primary 47H09, 90C25; Secondary 26A51, 26B25.

## 1 Introduction

We denote by $\mathcal{H}$ a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We say that an operator $T : \mathcal{H} \to \mathcal{H}$ is *convergent* if, for every starting vector $x^0$, the sequence $\{x^k\}$ defined by $x^k = Tx^{k-1}$ converges weakly to a fixed point of $T$, whenever $T$ has a fixed point. Fixed-point iterative methods are used to solve a variety of problems by selecting a convergent $T$ for which the fixed points of $T$ are solutions of the original problem. It is important, therefore, to identify properties of an operator $T$ that guarantee that $T$ is convergent.

An operator $T : \mathcal{H} \to \mathcal{H}$ is nonexpansive if, for all $x$ and $y$ in $\mathcal{H}$,

$$\|Tx - Ty\| \leq \|x - y\|. \tag{1.1}$$

*Journal of Convex Analysis, **22(4)**, pp. 963–967, 2015

†Charles_Byrne@uml.edu, Department of Mathematical Sciences, University of Massachusetts Lowell, Lowell, MA 01854

Just being nonexpansive does not make $T$ convergent, as the example $T = -Id$ shows; here $Id$ is the identity operator. It doesn't take much, however, to convert a nonexpansive operator $N$ into a convergent operator. Let $0 < \alpha < 1$ and $T = (1-\alpha)Id + \alpha N$; then $T$ is convergent. Such operators are called *averaged* [1, 3, 6] and are convergent as a consequence of the Krasnosel'skii-Mann Theorem [5].

A operator $T : \mathcal{H} \to \mathcal{H}$ is *firmly nonexpansive* if, for all $x$ and $y$ in $\mathcal{H}$,

$$\langle Tx - Ty, x - y\rangle \geq \|Tx - Ty\|^2. \tag{1.2}$$

It is not hard to show that $T$ is firmly nonexpansive if and only if $T = \frac{1}{2}(Id + N)$, for some nonexpansive operator $N$. Clearly, then, if $T$ is firmly nonexpansive, $T$ is averaged, and therefore $T$ is nonexpansive, and all firmly nonexpansive operators are convergent. Also, $T$ is firmly nonexpansive if and only if $G = Id - T$ is firmly nonexpansive. The Baillon–Haddad Theorem is the following.

**Theorem 1.1 (The Baillon–Haddad Theorem)** ([2], Corollaire 10]) *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and Gâteaux differentiable on $\mathcal{H}$, and its gradient operator $T = \nabla f$ nonexpansive. Then $f$ is Fréchet differentiable and $T$ is firmly nonexpansive.*

In [2] this theorem appears as a corollary of a more general theorem concerning $n$-cyclically monotone operators in normed vector space. In [4] Bauschke and Combettes generalize the Baillon–Haddad Theorem, giving several additional conditions equivalent to the two in Theorem 1.1. Their proofs are not elementary.

The Baillon–Haddad Theorem provides an important link between convex optimization and fixed-point iteration. If $g : \mathcal{H} \to \mathbb{R}$ is a Gâteaux differentiable convex function and its gradient is $L$-Lipschitz continuous, that is,

$$\|\nabla g(x) - \nabla g(y)\| \leq L\|x - y\|, \tag{1.3}$$

for all $x$ and $y$, then $g$ is Fréchet differentiable and the gradient operator of the function $f = \frac{1}{L}g$ is nonexpansive. By the Baillon–Haddad Theorem the gradient operator of $f$ is firmly nonexpansive. It follows that, for any $0 < \gamma < \frac{2}{L}$, the operator $Id - \gamma\nabla g$ is averaged, and therefore convergent. The class of averaged operators is closed to finite products, and $P_C$, the orthogonal projection onto a closed convex set $C$, is firmly nonexpansive. Therefore, the projected gradient-descent algorithm with the iterative step

$$x^{k+1} = P_C(x^k - \gamma\nabla g(x^k)) \tag{1.4}$$

converges weakly to a minimizer, over $C$, of the function $g$, whenever such minimizers exist.

In this note we present a short and elementary proof of the following theorem, using only fundamental properties of convex differentiable functions.

**Theorem 1.2** *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and Gâteaux differentiable. The following are equivalent:*

1. *the function $F(x) = \frac{1}{2}\|x\|^2 - f(x)$ is convex;*
2. *for all $x$ and $z$ we have*

$$\frac{1}{2}\|z-x\|^2 \geq D_f(z,x) \doteq f(z) - f(x) - \langle \nabla f(x), z-x \rangle \geq 0; \tag{1.5}$$

3. *the gradient operator $T = \nabla f$ is firmly nonexpansive;*
4. *the function $f$ is Fréchet differentiable and $T = \nabla f$ is nonexpansive.*

The proof of Theorem 1.1 given in [10] was reproduced in [7]. The proof given here for Theorem 1.2 is based on the one given for Theorem 1.1 in [8] and [9].

# 2 Proof of Theorem 1.2

Prove (2), assuming (1), that $F$ is convex. Since $F$ is convex, and $\nabla F(x) = x - \nabla f(x)$, we have

$$F(z) \geq F(x) + \langle \nabla F(x), z-x \rangle, \tag{2.1}$$

which is equivalent to

$$\frac{1}{2}\|z-x\|^2 \geq D_f(z,x). \tag{2.2}$$

Prove (3), assuming (2), that Equation (1.5) holds. Let $y \in \mathcal{H}$ be arbitrary and fixed. Let $d(x) = D_f(x,y)$. Then $d(x)$ is convex and $\nabla d(x) = \nabla f(x) - \nabla f(y)$. It is easily seen that $D_f(z,x) = D_d(z,x)$, so from (1.5) we have

$$\frac{1}{2}\|z-x\|^2 \geq D_d(z,x) = d(z) - d(x) - \langle \nabla f(x) - \nabla f(y), z-x \rangle. \tag{2.3}$$

Now let $z = x - \nabla f(x) + \nabla f(y)$. Inserting this $z$ into (2.3), we obtain

$$D_f(x,y) = d(x) \geq d(z) + \frac{1}{2}\|\nabla f(x) - \nabla f(y)\|^2 \geq \frac{1}{2}\|\nabla f(x) - \nabla f(y)\|^2. \tag{2.4}$$

Similarly, we can show that

$$D_f(y,x) \geq \frac{1}{2}\|\nabla f(x) - \nabla f(y)\|^2. \tag{2.5}$$

Adding the previous two inequalities, we get

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \|\nabla f(x) - \nabla f(y)\|^2, \tag{2.6}$$

so $T = \nabla f$ is firmly nonexpansive. Prove (4), assuming (3), that $T = \nabla f$ is firmly nonexpansive. It is clear that $T$ is then nonexpansive. Since the gradient operator is continuous, $f$ is Fréchet differentiable. Prove (1), assuming (4). The function $F(x) = \frac{1}{2}\|x\|^2 - f(x)$ is Fréchet differentiable and $\nabla F(x) = x - \nabla f(x)$. Since

$$\langle \nabla F(z) - \nabla F(x), z - x \rangle \geq \|z - x\|(\|z - x\| - \|\nabla f(z) - \nabla f(x)\|) \geq 0,$$

we know that $F(x)$ is a convex function. ∎

Notice that we have actually proved a somewhat stronger inequality than (1.2):

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle - \|\nabla f(x) - \nabla f(y)\|^2$$

$$\geq D_f(x - \nabla f(x) + \nabla f(y), y) + D_f(y - \nabla f(y) + \nabla f(x), x) \geq 0. \tag{2.7}$$

We get a slightly more general version of Theorem 1.2, but with a less elementary proof, if we assume that $f$ is lower semi-continuous and omit the assumption that $f$ is Gâteaux differentiable. Once we assume 1., the Gâteaux differentiablility of $f$ follows from Proposition 2.1 and Proposition 2.2.

**Definition 2.1** *Let $f : \mathcal{H} \to \mathbb{R}$ be arbitrary. The* subdifferential *of $f$, at the point $x$, is the set*

$$\partial f(x) = \{u | \langle u, y - x \rangle \leq f(y) - f(x), \text{for all}\, y \in \mathcal{H}\}.$$

*The members of $\partial f(x)$ are the* subgradients *of $f$ at $x$.*

For any lower semi-continuous $f : \mathcal{H} \to \mathbb{R}$, the function $f$ is convex if and only if $\partial f(x)$ is not empty, for all $x$ (see [5], Proposition 16.14).

**Proposition 2.1** *Let $f : \mathcal{H} \to \mathbb{R}$ and $g : \mathcal{H} \to \mathbb{R}$ be arbitrary functions. Then*

$$\partial f(x) + \partial g(x) \subseteq \partial(f + g)(x). \tag{2.8}$$

**Proof:** The containment follows immediately from the definition of the subdifferential. ∎

Since both $f$ and $g$ are finite-valued, the containment in (2.8) actually goes both ways, if both $f$ and $g$ are lower semi-continuous. In some discussions, convex functions may be allowed to take on the value $+\infty$. In such cases only the containment in (2.8) may hold. See Corollary 16.38 of [5].

**Proposition 2.2** *If $f : \mathcal{H} \to \mathbb{R}$ and $g : \mathcal{H} \to \mathbb{R}$ are both lower semi-continuous and convex, and $f + g = h$ is Gâteaux differentiable, then both $f$ and $g$ are Gâteaux differentiable.*

**Proof:** From Proposition 2.1 we have

$$\partial f(x) + \partial g(x) \subseteq \partial(f+g)(x) = \partial h(x) = \{\nabla h(x)\}.$$

Therefore, both $\partial f(x)$ and $\partial g(x)$ are nonempty and must be singleton sets. Therefore, both functions are Gâteaux differentiable, according to Proposition 17.26 of [5]. ∎

# References


1. Baillon, J.-B., Bruck, R.E., and Reich, S. (1978) "On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces." *Houston Journal of Mathematics*, **4**, pp. 1–9.

2. Baillon, J.-B., and Haddad, G. (1977) "Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones." *Israel J. of Mathematics*, **26**, pp. 137-150.

3. Bauschke, H., and Borwein, J. (1996) "On projection algorithms for solving convex feasibility problems." *SIAM Review*, **38 (3)**, pp. 367–426.

4. Bauschke, H., and Combettes, P. (2010) "The Baillon-Haddad Theorem Revisited." *J. Convex Analysis*, **17**, pp. 781–787.

5. Bauschke, H., and Combettes, P. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, New York: Springer CMS Books in Mathematics, 2011.

6. Byrne, C. (2004) "A unified treatment of some iterative algorithms in signal processing and image reconstruction."*Inverse Problems* **20**, pp. 103–120.

7. Byrne, C. (2007) *Applied Iterative Methods.* Wellesley, MA: A K Peters.

8. Byrne, C. (2014) *Iterative Optimization in Inverse Problems.* Boca Raton, FL: CRC Press.

9. Byrne, C. (2014) *A First Course in Optimization.* Boca Raton, FL: CRC Press.

10. Golshtein, E., and Tretyakov, N. (1996) *Modified Lagrangians and Monotone Maps in Optimization.* New York: John Wiley and Sons.